CHAPTER 1

General introduction

Kamila Z. Rosłoniec, Robert van der Geize and Lubbert Dijkhuizen

Department of Microbiology, Groningen Biomolecular Sciences and Biotechnology Institute (GBB), University of Groningen, P.O. Box 14, 9750 AA Haren, The Netherlands

To be submitted for publication

Actinomycetes

"Actinomycetes" is a term for a suprageneric taxon [56] comprised of Gram positive heterotrophic and morphologically diverse bacteria, dwelling in a wide variety of ecosystems. Actinomycetes often contain a high guanine (G) plus cytosine (C) content in their DNA and are recognized as environmentally, clinically and industrially important [315]. First regarded as Fungi, actinomycete bacteria have been classified to the order *Actinomycetales* (ray fungus). *Actinomycetales* belong to the class *Actinobacteria* and to the phylum *Actinobacteria*. The ancient and phylogenetically diverged phylum *Actinobacteria* constitutes the largest lineage among phyla of the *Bacteria* domain [315]. Actinomycetes are best known for comprising a number of pathogens (e.g. *Mycobacterium tuberculosis*), for strains that are able to produce bioactive compounds (e.g. *Streptomyces*, *Amycolatopsis*) and for strains forming mycelium and hyphae, morphologically resembling fungi (e.g. *Streptomyces* spp.).

Actinomycetes are detected and systematized on basis of their 16S rRNA gene analysis, DNA:DNA pairing studies, chemotaxonomic and phenotypic tests into suborders, families and genera, where especially the latter two taxonomic groups are constantly changing due to the growing number of new strains that are being discovered and taxonomically identified [79, 144, 279].

Actinomycetes are aerobic and microaerobic [61] and their cell walls may contain mycolic acids, defined as high-molecular weight β-hydroxy fatty acids, α-branched in two positions with long alkyl chains [144]. Mycolic acids facilitate the uptake of hydrophobic substrates into the cell [179], they contribute to bacterial adhesion properties [277] and they are responsible for the formation of a permeability barrier, preventing antibiotics and other molecules to enter the cell. Mycolic acids containing

actinomycetes, designated mycolata, are thus environmentally persistent [70]. The mycolata belong to the suborder *Corynebacterineae,* which to date is comprised of 6 families, i.e. *Corynebacteriaceae*, *Dietziaceae*, *Segniliparaceae, Tsukamurellaceae, Mycobacteriaceae* and *Nocardiaceae*. The latter includes the genus *Rhodococcus* [338].

The genus *Rhodococcus*

Rhodococci are aerobic, non-motile, non-sporulating, lysozyme sensitive mycolata, with DNA containing 63-73 mol% of G+C [79]. The genus *Rhodococcus* was first proposed for red-pigmented bacteria by Zopf in 1891 and since then the genus systematics changed considerably [79]. Presently, this genus is paraphyletic, i.e. the genus comprises some, but not all of the descendants from a common ancestor [79], which is indicative for necessary future taxonomic improvements associated with identification of new species [84].

Ecology

Rhodococcal species are isolated mainly from soil, but they are also common in other environments, like in water, activated sludge foams, guts of blood-sucking arthropods, herbivore dung [79], healthy and sick animals, rocks [14] and deserts [96]. They encompass symbionts (e.g. *Rhodococcus rhodnii*) and pathogens of plants (e.g. *Rhodococcus fascians*) and animals, including humans (e.g. *Rhodococcus equi*) [14].

Pleomorphism

Rhodococci display morphological heterogeneity, which is dependent on the strain, growth conditions and phase of growth [161], but all have a short rod or a coccoidal form at the start of their life cycle [79]. Rods and cocci may further germinate. Filaments,

small branches, aerial synnemata built from unbranched filaments, microscopically small, feeble aerial hyphae or even branched hyphae can be formed. A new generation begins with fragmentation [79]. Interestingly, it was recently reported that several *Rhodococcus* strains form specialized aerial architectures, resembling miniature mushrooms, for utilization of vaporized alkyl phenols, such as *p*-cresol, as the sole carbon and energy source [314].

Cell envelope

The rhodococcal cell envelope skeleton is composed of three main layers, located over the cell membrane of phospholipids and proteins [284]. The first innermost layer is of peptidoglycan type $A1_{\gamma}$, containing *meso*-diaminopimelic (*meso*-A_{2pm}) acid [56, 284] and muramic acid with N-glycolyl residues ("A" stands for cross-linkage between peptide subunits at position 3 and 4; "1" for absence of the peptide bridge; "$_{\gamma}$" for the presence of the *meso*-A_{2pm}) [79]. The second layer is of arabinogalactan, covalently linked with peptidoglycan [56] and with a large amount of mycolic acids [246]. Mycolic acids contain 30-54 carbon atoms [79] and form the third layer of the cell envelope, functioning as an outer permeability barrier [246]. Mycolic acids probably also occur in a free [284] and in a partially free form [277]; in the latter case they can be esterified by trehalose (α-D-glucopyranosyl-α-D-glucopyranoside) [294]. In *Rhodococcus* spp. mycolic acids form up to 40% of the cell envelope skeleton [277].

Environmental conditions, such as culture media composition, induce changes in rhodococcal cell wall lipids, to maintain cell wall fluidity and impermeability. Moreover, cell envelope lipid profiles of *Rhodococcus* strains are known to be dependent on the phase of growth [277].

The outer membrane of some *Rhodococcus* strains, e.g. *Rhodococcus erythropolis*, contains porin forming channels [84] for hydrophilic solutes [317]. At least two types of channels, cation and anion selective ones, have been proposed for *R. equi* [246]. Interestingly, the cell of this pathogen is enclosed in a thick, lamellar and antigenically variable capsule of polysaccharide [317]. The specific capsular polysaccharide, produced by *R. equi* serotype 4, was found to be a high-molecular-weight acidic polymer, composed of D-glucose, D-mannose, pyruvic acid and 5-amino-3,5-dideoxynonulosonic (rhodaminic) acid, in a 2:1:1:1 ratio [262]. The capsule is present in virulent and avirulent isolates of *R. equi* and probably is not a virulence factor, but may provide protection against desiccation [286].

Metabolic properties

Rhodococci are catalase-positive, usually partially acid-alcohol fast and they possess an oxidative carbohydrate metabolism [79]. They utilize a wide variety of organic compounds as sole carbon and energy source, including gaseous hydrocarbons [14] and xenobiotics, such as aliphatic, aromatic and chlorinated hydrocarbons, oxygenates, nitroaromatics, nitriles, halogenated and heterocyclic compounds [179], under diverse conditions [44]. Many pollutants that are degraded by *Rhodococcus* sp. are toxic and recalcitrant, such as polychlorinated biphenyls, present in paints, plastics, insulating fluids [179], sulphonated azo dyes [14], used for textile coloring, or hexahydro-1,3,5-trinitro-1,3,5-triazine, a widely used explosive [260], to mention a few examples.

Rhodococci may be used for degradation of waste materials, ideally leading to production of useful compounds. It has been shown recently that *Rhodococcus aetherivorans* strain IAR1 is able to use toluene, a volatile pollutant, as the sole carbon source,

with concomitant production of biodegradable plastic that is poly(3-hydroxybutyrate-*co*-3-hydroxyvalerate) [103].

The impressive catabolic capabilities of rhodococci have been examined, and even improved upon in various studies. The best known example is that of industrial acrylamide production by *Rhodococcus rhodochrous* strain J1 [110]. Moreover, it has been shown recently that *R. erythropolis* strain IGST8 decorated with magnetic Fe_3O_4 nanoparticles performed better than a non-decorated strain in desulphurization of dibenzothiopene (DBT) [8]. DBT together with methylated DBTs contribute up to 70% to the sulphur pool present in fossil fuels [8]. DBT is particularly recalcitrant to hydrodesulphurization, an industrial pre-combustion method that concomitantly generates sulphur oxides, contributing to air pollution and acid rain [151, 235, 301].

Being metabolic marvels, rhodococci are a source of novel enzymes [79]. A unique endoglycoceramidase has been described of *R. equi*, hydrolyzing the β-galactosidic linkage of oligo-galactosylceramides (6-gala GSLs, 6-gala series glycosphingolipids) into intact galacto-oligosaccharides and ceramides, as well as catalyzing a transglycosylation reaction. The transglycosylation activity of the enzyme was used for developing a new method for detecting 6-gala GSLs, further facilitating research into these compounds, of which the function is still not fully known [112]. Bifunctional activity was also presented for nitrile hydratase from *Rhodococcus* sp. strain N771: apart from the hydration of nitriles to the corresponding amides, it also catalyzed conversion of isobutylisonitrile into isobutylamine [290].

Transformations of nitrile as well as e.g. conversions of racemic α-substituted phenylacetonitriles and cyclopropanecarbonitriles can be conducted by rhodococci in enantio-selective manner yielding enantiopure products, such as carboxylic acids

and amides [83]. *R. erythropolis* strain AJ270 was demonstrated to conduct conversions of this type in very high yields [166].

In addition to production of the above mentioned polymers and amides, rhodococci are also able to synthesize exopolysaccharides [297, 298], flocculants and surfactants [14]. Furthermore, *R. erythropolis* strain JCM 6824 was demonstrated to produce aurachin RE, a quinoline antibiotic, exhibiting a wide and strong antimicrobial spectrum against Gram positive bacteria [140]. *Rhodococcus* sp. strain K01-B0171 produces antimycobacterial cyclic peptides, lariatins A and B [12]. Rhodococci are also able to accumulate storage compounds, likely in response to environmental stresses. Triacylglycerols, wax esters, polyphosphate, glycogen as well as co-polyesters containing 3-hydroxybutyric acid and 3-hydoxyvaleric acid are storage compounds of *Rhodococcus jostii* strain RHA1 [96].

Production of biosurfactants

Biosurfactants are synthesized by several *Rhodococcus* species (e.g. *R. erythropolis*) [14], in the presence of *n*-alkanes in liquid culture media [160]. Biosurfactants formed by *Rhodococcus* sp. strain SD-74 from *n*-hexadecane, under high osmotic conditions, are mainly succinoyl trehalose lipids [293]. Production of biosurfactants and changes in physicochemical properties of the cell surface are the most distinct auto-ecological adaptations for rhodococci, in response to the presence of *n*-alkane hydrocarbons [44].

Rhodococcal biosurfactants are biochemically versatile [293] and they effectively reduce surface as well as interfacial tensions of water and aqueous solutions [177]. They are furthermore of lower toxicity and of higher biodegradability than a number of synthetic surfactants [225]. Rhodococcal biosurfactants are thus interesting for industry, as pharmaceuticals, cosmetics, in food, for high-tech applications, e.g. in electronic

printing and also for bioremediation [177, 225]. Hence, efforts are being currently made to increase the fermentative yield of these surface-active metabolites [225], to elucidate their structure [177, 293], to characterize their physicochemical properties [44, 177, 293] and cellular toxicity [177].

Bioremediation

The impressive catabolic abilities of rhodococci, together with their environmental persistence and adaptation, using high affinity uptake systems [192, 194], accumulation of storage compounds, ability to adhere to hydrophobic substrates and to produce surfactants, make them good candidates for application in bioremediation [162]. The presence of other, more easily metabolizable chemicals does not always repress degradation of pollutants by *Rhodococcus* strains [14]. Immobilized cells of *Rhodococcus* sp. strain F92 on polyurethane foam degrade various petroleum products [243]. Trichloroethylene degradation by *Rhodococcus* sp. strain L4 was found to be induced by cumin, lemon and lemon grass oils instead of toluene [285]. Efforts are being made to apply *Rhodococcus* sp. strain EH831, immobilized on polyurethane biofilter, for removal of benzene and toluene mixture under transient loading, resembling real working conditions of industrial biofilters [165].

Rhodococci also are able to accumulate heavy metals, including radioactive ions, such as high levels of cesium (^{137}Cs) [14], with good opportunities for application in bioremediation.

Xenobiotic interactions, however, may have a strong effect on the catabolic abilities of rhodococcal strains, as was demonstrated for *Rhodococcus* sp. strain EH831. This strain was able to degrade benzene and methyl tert-butyl ether when supplied as a single component, but was inhibited in degradation of these compounds when subjected to a

mixture of them, additionally containing ethylbenzene [164]. Rhodococci also are known to contribute in the formation of problematic surface foams in activated sludge plants [45]. These foams, which are easily dispersed by wind, reduce oxygen transfer, leading to an increase in effluent demand for oxygen and to suspension of solids (also pathogens) in effluents. Application of *Rhodococcus* strains in bioremediation thus is not easy, the more because there is still a limited understanding of rhodococcal catabolic enzyme functions, regulation of gene expression, and the role of global control mechanisms [179].

Genomes and plasmids of Rhodococcus strains

The metabolic versatility of *Rhodococcus* strains is reflected in their genomes [301], encoding numerous catabolic pathways for a variety of chemical compounds that subsequently are funneled into fewer "central" pathways [236]. Metabolic pathways in rhodococci are complex, often carefully regulated, and at the same time flexible, due to genomic rearrangements [161]. In order to better understand the genetics of rhodococci, availability and analysis of their genomic sequences is needed. The complete genome sequences of *R. jostii* RHA1 (9.7 Mb; www.rhodococcus.ca; GenBank CP000431), *R. opacus* B4 (7.9 Mb, without plasmids; GenBank AP011115), *R. erythropolis* PR4 (6.9 Mb; GenBank AP008957), *R. erythropolis* SK121 (6.8 Mb, NZ_ACNO00000000), *R. equi* 103S (5.1 Mb; http://www.sanger.ac.uk/Projects/R_equi/) and partial genome sequences of *R. equi* ATCC 33707, *R. erythropolis* DSM8424, *R. opacus* PD630, *Rhodococcus* sp. NS1, *Rhodococcus* sp. B264-1, *R. aetherivorans* and *R. rhodochrous* are known. Chromosomal and plasmid DNA of *Rhodococcus* strains is either linear or circular. Both *R. jostii* RHA1 and *R. opacus* B4 have linear chromosomes, while *R. erythropolis* PR4 has a circular one. One strain may have both linear and circular plasmids that generally evolve more rapidly than the chromosome. For example,

R. opacus B4 has three circular plasmids (pKNR [110 kb; NC_012523], pKNR01 [4.4 kb; NC_006969], and pKNR02 [2.8 kb; NC_006970]) and two linear plasmids (pROB01 [560 kb; NC_012520] and pROB02 [240 kb; NC_012521]). Some rhodococcal plasmids are known for their ability to conjugate [329].

Different replicons present in one cell may have diverse origins and they may stabilize each other, as was suggested for *R. jostii* RHA1 [189]. On the other hand, some rhodococci, e.g. *R. rhodochrous* strain NCIMB13064 [148], are known for their genomic instability (reviewed by [161]). *R. fascians* is able to integrate plasmids in its genome at random sequences [54] and different strains can recombine (heterothallism), indicating flexibility of rhodococcal genomes (reviewed by [161]). The genome of *R. jostii* RHA1 has probably evolved by ancient gene acquisition or gene duplication and is more stable [189]. Rhodococci are known for possessing highly similar genes encoding homologous enzymes [301], potentially leading to the simultaneous use of compounds present in their environment [189]. Transcriptional regulation appears to be abundant in rhodococci, as evidenced by the presence of genes encoding regulatory proteins and by the dependence of gene transcription on inducers, repressors and other gene expression regulation factors [25, 111, 132, 145, 288, 304, 309].

Genetic engineering of Rhodococcus strains

With more and more sequence data available, further exploration and engineering of *Rhodococcus* strains is possible, employing the rhodococcal genetic tools that have been developed in recent years. Many *Rhodococcus*-*E. coli* shuttle vectors have been constructed based on rhodococcal cryptic plasmids, for studies on gene expression, gene function and for construction of genomic libraries [19, 46, 59, 91, 98, 146, 169, 185, 186, 205, 273, 306]. Efficient rhodococcal expression systems, e.g. pTip and pNit vectors

[205] and transposon-based vectors [251] have greatly facilitated production of enzymes by *Rhodococcus* strains. Several rhodococcal reporter systems have been successfully used in gene expression studies [143, 161, 304]. There are strategies known for the construction of random and targeted mutations in rhodococci [300, 301, 305, 317], enabling the functional analysis of genes and enzymes. The unmarked in-frame gene deletion strategy is particularly useful, because it allows construction of multiple targeted deletions and maintenance of the mutated *Rhodococcus* strains in the absence of any antibiotics [300, 305].

Genetic engineering lies at the basis of metabolic engineering, aiming to construct better performing producer strains, especially for synthesis of desired metabolites in higher yields, via manipulation of fluxes in a metabolic network [7]. Interestingly, metabolically engineered *Rhodococcus* strains converting indene to *trans*-(1*R*,2*R*)-indandiol, a precursor for the AIDS drug (Crixivan), have been constructed successfully [280]. In recent years our research group has focused on the construction of *Rhodococcus* strains converting cheap phytosterols or cholesterol into bioactive steroids (see below for more details).

Steroids

Steroids are bioactive organic compounds, necessary for the functioning of eukaryotic organisms. Steroidal molecules encompass sterols, many hormones, bile acids, vitamin D and various secondary metabolites, e.g. saponines [18]. Steroids all have a cyclopenta[α]phenantrene skeleton (Fig. 1a), either partially or completely hydrogenated. Various derivatives exist, harboring one or more ring expansions,

contractions or bond scissions. Steroids with a hydroxyl group at carbon atom 3 (C3) are defined as sterols (Fig. 1b), [118].

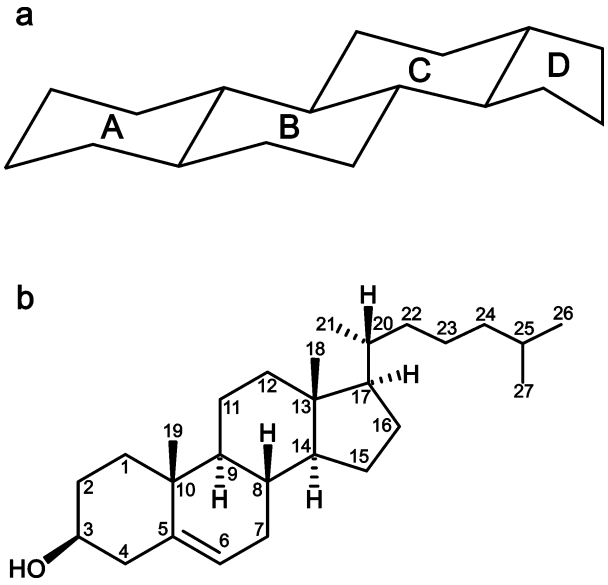


Figure 1. Formula of (a) the stereochemical cyclopenta-[α]-phenanthrene skeleton of steroids, depicted as a reduced ring system in the *trans*-form composed of three fused cyclohexane rings (A, B and C) in a phenanthrene arrangement and a terminal cyclopentane ring (D) (b) (3*S*,8*S*,9*S*,10*R*,13R,14*S*,17*R*)-17-[(1*R*)-1,5-dimethylhexyl]-10,13-dimethyl-2,3,4,7,8,9,11,12,14,15,16,17-dodecahydro-1*H*-cyclopenta[α]phenanthren-3-ol, better known as cholesterol (5-cholestene-3β-ol), one of the most abundant natural steroids, with carbon atom numbering.

Sterols yield fluidity and flexibility to eukaryotic membranes [204]. The sterol biosynthetic pathway has an essential role in the metabolism of eukaryotes [47]. In mammals, insects and photosynthetic plants, sterols are also transformed into steroidal hormones [217]. Mineralocorticoids, glucocorticoids and sex hormones function as transcription factors regulating gene expression, by binding to and activating receptor molecules, and thus are crucial for controlling biological processes in animals and humans [23]. Sterols of animals play a role as parahormones, autohormones, as agents for intercellular communication, as transducers of intracellular signals and as reproductive regulators [176].

Prokaryotes generally do not synthesize steroids. Their cell membranes may contain sterol-like compounds, pentacyclic terpenes of the hopanoid type [20], which are

synthesized from squalene [204]. Only few bacterial species were found to produce steroids and their precursors, namely myxobacteria such as *Stigmatella aurantiaca* [20], methylococci such as the methanotrophic bacterium *Methylococcus capsulatus* strain Bath [156, 204], members of the genus *Mycoplasma*, e.g. *M. capricolum* [40], the planctomycete *Gemmata obscuriglobus* [230], and the actinomycetes *Gordonia australis* strain Acta 2299 [255] as well as *Mycobacterium smegmatis* [157].

G. australis strain Acta 2299 synthesizes novel steroids: bendigole A, B and C, among which A and C display androgenic activities [255]. *M. smegmatis* was shown to produce cholesterol [157]. No other actinobacterium was demonstrated to synthesize sterols despite possessing putative sterol biosynthetic genes, such as sterol 14α-demethylase gene (*CYP51*). Desmond *et al*. [52] discuss in detail bacterial sterol biosynthetic genes and suggest acquisition of these genes from eukaryotes via horizontal gene transfer. The roles of steroids/sterols in bacteria remain to be elucidated.

Also, bile acids are synthesized by marine bacterial strains, such as *Streptomycetes* (e.g. *S. faecium* strain AN-21), *Myroides* (e.g. *Myroides* sp. strain SM1) and *Rhodococcus* (e.g. *R. marinonascens* strain DSM 43752) [135].

In vertebrates and in microorganisms such as fungi, trypanosomatids and yeasts, steroids are formed from lanosterol, which is produced by enzymatic conversion of squalene, synthesized from mevalonate via the isoprenoid pathway [47]. Cycloartenol is a steroid precursor of higher plants, originating from oxidosqualene cyclization [217]. Since enzymatic cyclizations of (oxido)squalene are biochemically very complex processes, they are being investigated extensively and utilized for synthesis of novel triterpenes [1, 2, 34].

Interestingly, the plant species *Arabidopsis thaliana* was recently shown to possess two biosynthetic pathways for phytosterols (plant sterols), such as campesterol, sitosterol and stigmasterol, namely the cycloartenol and the lanosterol pathways [217]. Sterol and steroid biosynthesis is greatly dependent on reactions catalyzed by cytochrome P450 monooxygenase enzymes (CYPs). For instance, lanosterol, a precursor of animal cholesterol and fungal ergosterol, as well as obtusifoliol, a phytosterol precursor, are first transformed by CYP51. CYP51, present in organisms from all kingdoms of life, catalyzes in three steps the formation of 14α-demethylated products [167].

Phytosterols, such as β-sitosterol, as well as plant diosgenin (a steroid sapogenin) are used by the pharmaceutical industry for production of steroidal drugs. The importance of steroidal hormones as drugs for health treatments and the useful properties of steroid hormone derivatives in anabolic, anti-inflammatory, antirheumatic, contraceptive or sedative pharmaceuticals, made steroidal drugs highly attractive products [23]. Plant sterols are also utilized by several other sectors of industry, namely in food, for manufacturing of anti-cholesterol and anti-cancer additives of functional foods and in cosmetics, for making creams and lipsticks [63].

Exploitation of plants for the isolation of natural steroid compounds has its environmental limitations. Chemical synthesis of steroids is often laborious, or not feasible due to process complexity and low yields. In recent years, efficient and cost-effective steroid bioconversion processes, involving engineered microorganisms and their enzymes, have emerged, providing promising alternatives for chemical production and modification of steroid compounds (see below for more details).

Steroid metabolism in *Rhodococcus* strains

Rhodococci are unable to synthesize steroids *de novo*. However, they grow on sterols/steroids as a sole carbon- and energy source. Phytosterols, such as β-sitosterol (5-cholestene-24β-ethyl-3β-ol), β-sitostanol (5α-cholestan-24β-ethyl-3β-ol) and campesterol (5-cholestene-24α-methyl-3β-ol), as well as animal sterols, such as cholesterol (5-cholestene-3β-ol), epicholesterol (5-cholestene-3α-ol) and 5α-cholestanol (5α-cholestane-3β-ol), are degraded by rhodococci [250], (Chapter 2). Possessing a full set of catabolic enzymes, rhodococci are particularly interesting organisms for steroid production via bioconversion and biotransformation. Degradation of low-cost sterols from plant and animal waste material, by rhodococci, yielding various intermediates that may be harvested as bioactive steroids have great commercial potential [4, 119, 303, 308]. *Rhodococcus* strains are relatively easy to cultivate, they are mostly nonpathogenic, their genomic sequences are being revealed, and the molecular toolbox for rhodococcal gene manipulation and functional investigation is steadily increasing.

To apply rhodococci in sterol/steroid biotransformation processes, the enzymes involved in opening of the steroid polycyclic ring need to be inactivated [301]. Mutants of *Rhodococcus* strains blocked in various steroid degradation steps have been constructed, resulting in accumulation, from sterols, of bioactive steroids, which may be used as precursors in the synthesis of steroidal drugs [302, 303, 306, 308]. Investigation of these *Rhodococcus* mutant strains and characterization of enzymes involved in steroid catabolism has revealed a complex steroid degradation pathway and provided a better understanding of steroid metabolism [142, 233, 250, 301, 304, 309], (Chapter 2). Such thorough knowledge on steroid catabolism will facilitate the advanced engineering of production strains.

The microbial degradation of sterols involves three processes: sterol uptake, steroid side chain elimination and steroid ring opening [301]. The order of occurrence of steroid side-chain oxidation and steroid ring structure opening appears to depend on the rhodococcal strain. Transcriptome analysis of cholesterol grown cells of *R. jostii* strain RHA1 identified 572 upregulated genes including 6 gene clusters. Two of the strain RHA1 gene clusters appeared, based on annotation, to have a likely role in the following processes: one gene cluster in cholesterol degradation to 9,17-dioxo-1,2,3,4,10,19-hexanorandrostan-5-oic acid (DOHNAA) and the other gene cluster putatively in opening of DOHNAA steroid ring D (Fig. 2, 3). The function of chosen proteins encoded by genes within the cholesterol catabolic gene cluster was experimentally confirmed. The cholesterol catabolic cluster, comprised of 51 genes in *R. jostii* strain RHA1, was also found as highly conserved in mycobacterial strains, namely *M. tuberculosis* H37Rv, *Mycobacterium bovis* bacillus Calmette-Guérin and *Mycobacterium avium* (*subsp. paratuberculosis*) [309].

Steroid uptake is encoded by the mce4 gene cluster

Steroid transport into the cell is one of the key steps in steroid degradation. *R. jostii* strain RHA1 appears to require a set of proteins for the functional uptake of steroids through the thick cell wall, including an ATPase (Mce4G), the SupA and SupB permease subunits and the Mce4A-Mce4I proteins, together believed to constitute the Mce4 steroid uptake system [194]. Mutational analysis of the *mce4* gene cluster has shown that genes encoding Mce4 proteins are involved in cholesterol catabolism [226, 309]. Mutant strains of RHA1, with inactivated *supAB*, *mce4ABCDEF* or *mce4HI* genes, were unable to grow on cholesterol, β-sitosterol, 5α-cholestanol or 5α-cholestanone. Growth on 4-androstene-3,17-dione, cholic acid and progesterone was unaffected, indicating that additional steroid

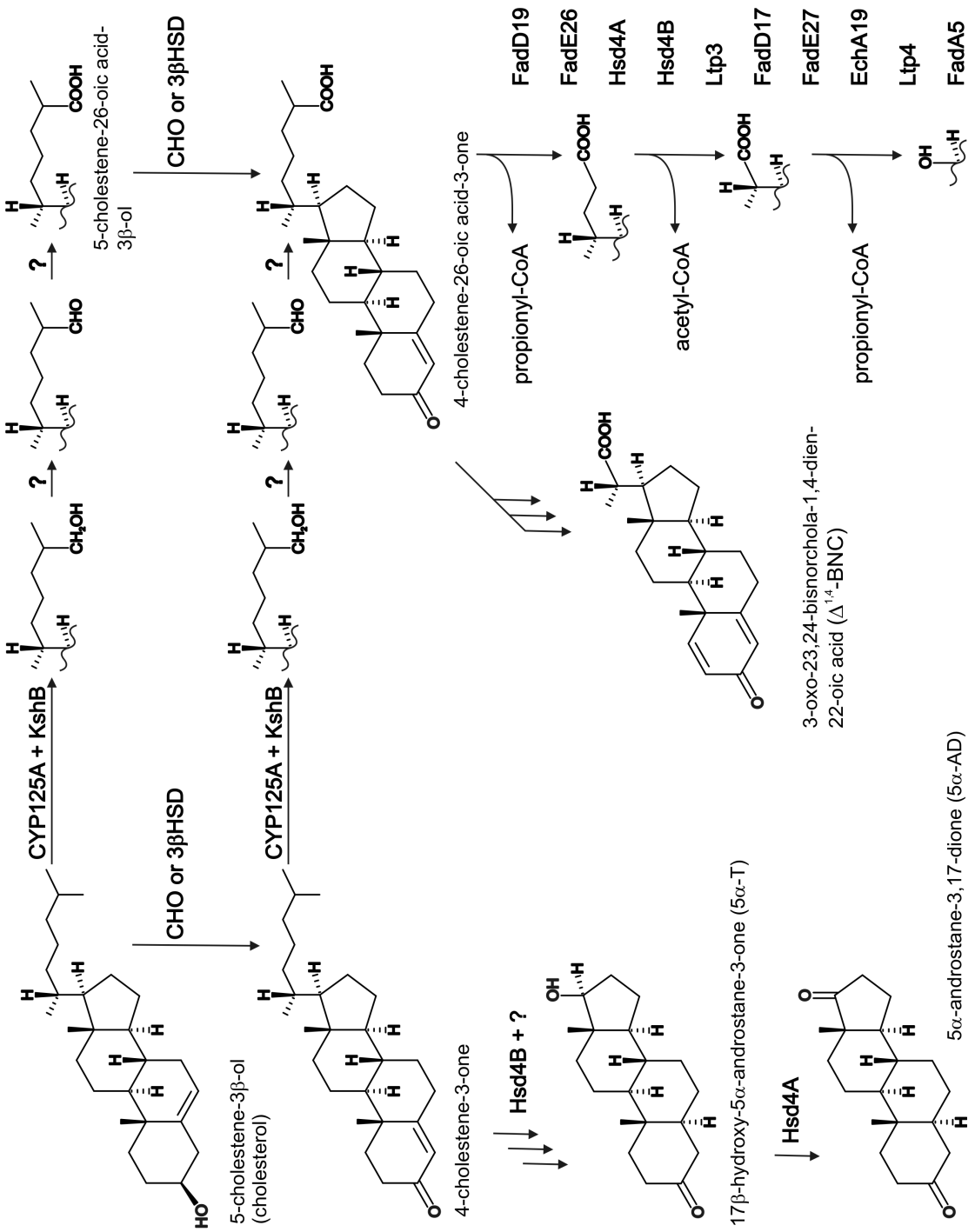


Figure 2 (panel A). Scheme of the cholesterol catabolic pathways in *Rhodococcus* (see text). Three arrows illustrate predicted multiple enzymatic steps. Reactions catalyzed by not yet identified enzymes are depicted with "?".

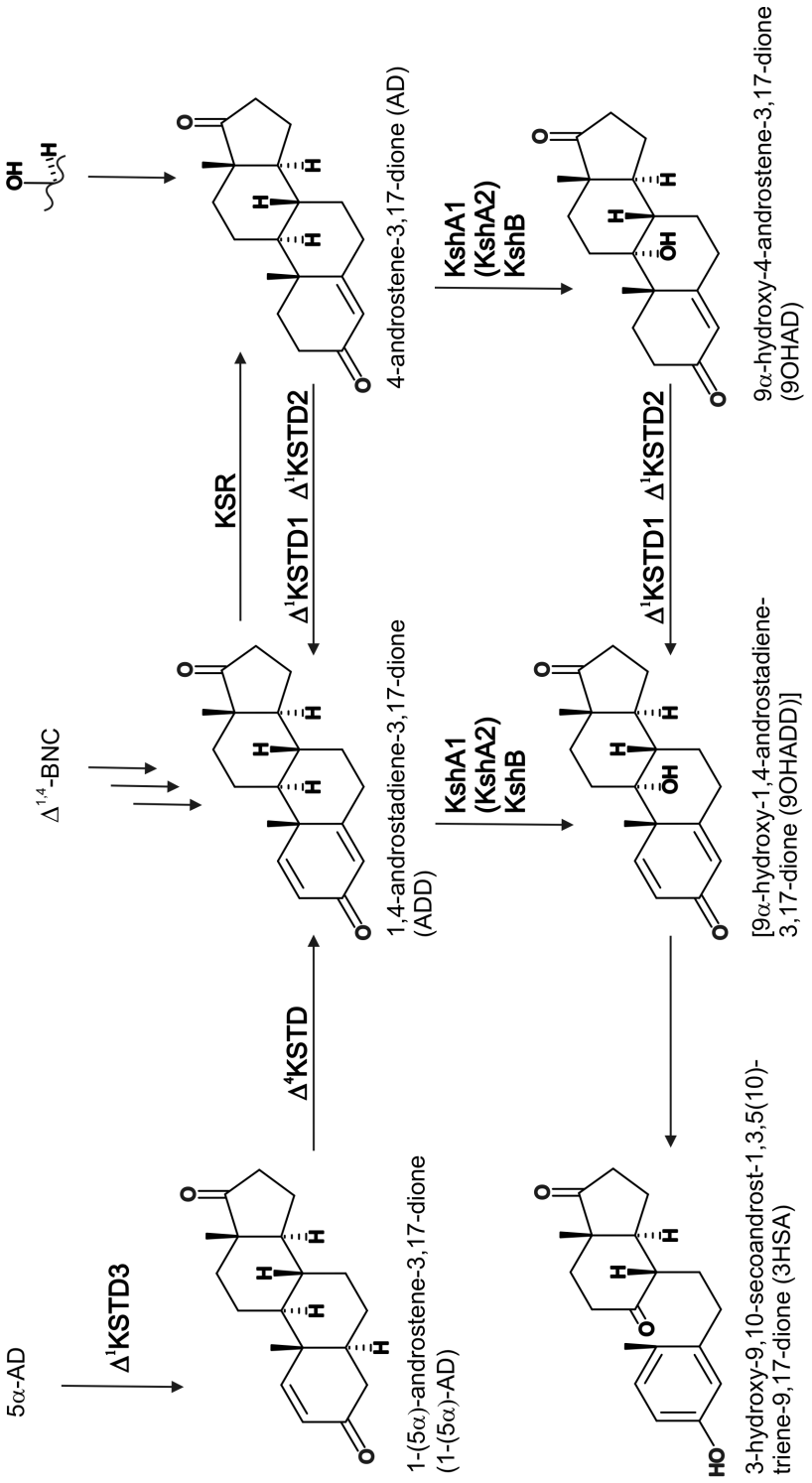


Figure 2 (panel B). Scheme of the cholesterol catabolic pathways in *Rhodococcus* (continuation of panel A; see text). Three arrows illustrate predicted multiple enzymatic steps.

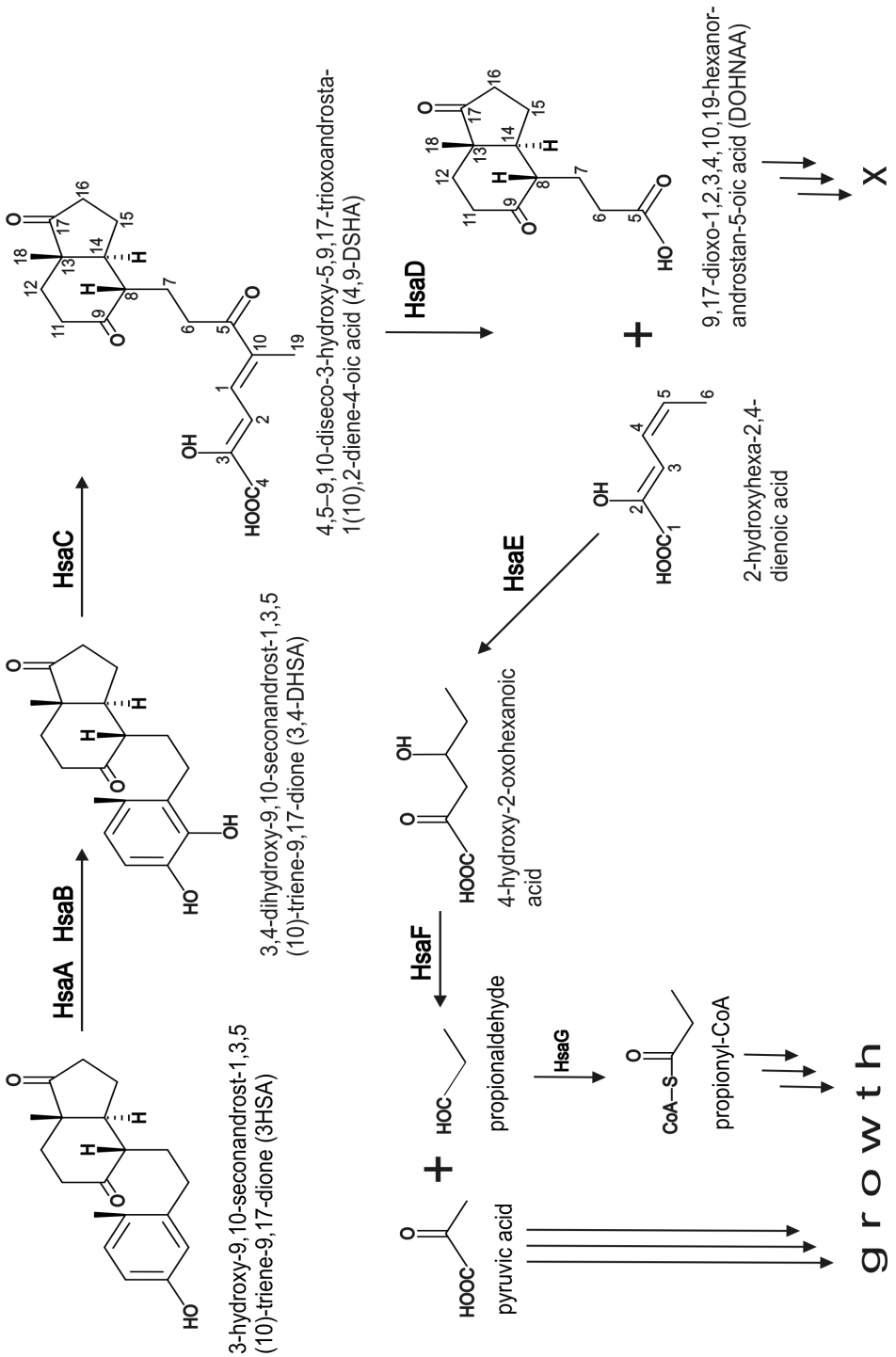


Figure 3. The 3HSA catabolic pathway (see text) with carbon atom numbering for 4,9-DSHA, DOHNAA and 2-hydroxyhexa-2,4-dienoic acid. Three arrows illustrate multiple enzymatic steps. The "x" depicts unknown compound(s).

uptake systems exist [194]. Signal sequences are predicted for most of the Mce4 proteins, indicating secretion or surface-exposition of these proteins in strain RHA1 [309]. The formation of a substrate-binding complex by Mce proteins that could mediate substrate movement across the cell wall of strain RHA1 has been proposed [194]. Detailed knowledge on the steroid uptake system of rhodococci is currently lacking, but will be important for both fundamental and applied research.

Steroid side chain degradation by Rhodococcus strains

Cleavage of the steroid side chain by *R. equi* is initiated by hydroxylation at C26 and oxidation to its corresponding carboxylic acid [202, 203]. The enzymology of steroid side chain cleavage by rhodococci is gradually elucidated. In strain RHA1, steroid side chain degradation is initiated by the cytochrome P450 enzyme CYP125(A14), catalyzing the steroid 26-hydroxylation of cholesterol and 4-cholesten-3-one (Fig. 2) [250], Chapter 2. CYP125(A14) was shown to be essential for growth of *R. jostii* RHA1 on 3-hydroxysterols. However, it remains to be determined whether complete side chain oxidation to cholestene-26-oic acid is catalyzed by CYP125(A14), or that a different enzyme is involved (Fig. 2) [250], (Chapter 2). The ortholog of CYP125(A14) in *M. tuberculosis* H37Rv ($CYP125A_{H37Rv}$) was shown to display steroid 26-hydroxylation activity towards cholesterol and 4-cholesten-3-one, initially attacking C26, C27 or both [27]. Selective C27-hydroxylation of cholesterol (5-cholestene-3β-ol) by $CYP125A_{H37Rv}$ has recently been demonstrated by [188].

The proposed mechanism for side-chain degradation of the C26-oic acid steroids in strain RHA1 is that of gradual shortening via a mechanism similar to β-oxidation of fatty acids [309] (Fig. 2). Support for this has recently been obtained by demonstrating that β-ketoacyl-coenzymeA (β-ketoacyl-CoA) thiolase activity (FadA5) of *M. tuberculosis*

H37Rv is necessary for the production of 4-androstene-3,17-dione (AD) and 1,4-androstadiene-3,17-dione (ADD) from cholesterol (Fig. 2) [213]. In strain RHA1, the orthologous gene encoding FadA5 (*fadA5*$_{RHA1}$) is located adjacent to the gene encoding CYP125(A14). Other putative thiolases of strain RHA1, namely Ltp3 and Ltp4, were proposed to play a role in cholesterol side chain β-oxidation, together with fatty acid-CoA ligase (FadD19), acyl-CoA dehydrogenase (FadE26), 17β-hydroxysteroid dehydrogenase (Hsd4A), 2-enoyl acyl-CoA hydratase (Hsd4B), fatty acid-CoA synthetase (FadD17), acyl-CoA dehydrogenase (FadE27) and fatty acid-CoA hydratase (EchA19) (Fig. 2) [309]. Depending on the strain of *Rhodococcus*, steroid side chain degradation may occur at various points in the sterol degradation pathway. Side chain degradation may be initiated as the first step in sterol degradation, or may occur only after steroid ring degradation has been initiated (Fig. 2) [250], (Chapter 2). In *R. jostii* strain RHA1, cholesterol side chain degradation and steroid ring catabolism both take place only after steroid C26-hydroxylation by CYP125(A14) has been performed. CYP125(A14) of strain RHA1 was shown to catalyze the initial step in sterol degradation and to constitute an essential step before ring oxidation could occur. On the contrary, inactivation of *CYP125A* in a mutant strain of *R. rhodochrous* strain DSM43269 did not affect ring oxidation and conversion of cholesterol into 4-cholesten-3-one and 1,4-cholestadiene-3-one was observed, indicating that ring oxidation is not dependent on side chain degradation [250], (Chapter 2).

Detailed information on rhodococcal sterol side chain cleavage will further facilitate construction of engineered *Rhodococcus* strains for the efficient production of steroid intermediates from sterols [309].

Steroid ring degradation by rhodococci

Rhodococcus strains generally produce several (iso)enzymes responsible for steroid ring opening [301]. The initial steps in steroid B-ring opening are catalyzed by 3-ketosteroid 9α-hydroxylase (KSH) and 3-ketosteroid Δ^1-dehydrogenase (Δ^1 KSTD) enzymes (Fig. 2) [301]. Subsequently, steroid ring A is aromatized and opened via *meta*-cleavage. The Hsa enzymes are involved in steroid ring A catabolism of 3-hydroxy-9,10-seconandrost-1,3,5(10)-triene-9,17-dione (3HSA) (Fig. 2 and 3) [58, 104, 309]. The Hsa pathway in strain RHA1 is composed of the following (putative) enzymes: 3HSA hydroxylase oxygenase (HsaA), 3HSA hydroxylase reductase (HsaB), 3,4-dihydroxy-9,10-seconandrost-1,3,5(10)-triene-9,17-dione (3,4-DHSA) dioxygenase (HsaC), 4,5-9,10-diseco-3-hydroxy-5,9,17-trioxoandrosta-1(10),2-diene-4-oic acid (4,9-DSHA) hydrolase (HsaD), 2-hydroxypentadienoate hydratase (HsaE), 4-hydroxy-2-oxovalerate aldolase (HsaF) and acetylaldehyde dehydrogenase (HsaG) (Fig. 2 and 3) [309]. Strain RHA1 has four predicted distinct steroid degradation pathways suggested to act specifically on different steroids. Each of these pathways appears to contain homologues of KshA, KSTD, as well as HsaA, HsaB, HsaC and HsaD [309].

Oxidation of 3β-hydroxysteroids

Cholesterol oxidase (CHO; EC 1.1.3.6) and 3β-hydroxysteroid dehydrogenase (3βHSD) both are able to catalyze 3β-hydroxysteroid oxidation, followed by (spontaneous) isomerization of Δ^5-3-ketosteroids into Δ^4-3-ketosteroids (Fig. 2) [55, 57, 330]. CHO as well as 3βHSD, characterized from *M. tuberculosis* H37Rv ($3\beta HSD_{H37Rv}$) and *Nocardia sp.*, both catalyze cholesterol transformation [57, 330]. In addition, dehydroepiandrosterone and 5-pregnen-3β-ol-20-one (pregnenolone) are known to be converted by $3\beta HSD_{H37Rv}$, whereas bacterial CHO has generally a larger substrate range,

i.e. 5α-cholestan-5-en-3β-ol (β-cholestanol), sitost-5-en-3β-ol (β-sitosterol), stigmast-5-en-3β-ol (β-stigmasterol), ergosta-5,7,22-trien-3β-ol (ergosterol), pregnenolone, 3β-hydroxyandrost-5-en-17-one (dehydroepiandrosterone), 5α-androstan-3-ol-17-one (epiandrosterone) [57].

Cholesterol oxidases of rhodococci and other actinobacteria have several biological functions in, for example, sterol catabolism and in biosynthesis of a macrolide, antifungal primaricin (*Streptomyces natalensis*). Considering the former role, CHO is generally regarded as the compulsory enzyme involved in the first step of bacterial sterol catabolism. Rhodococcal CHO, existing both as extracellular and cell-associated protein, is known to be induced by plant or animal sterols, or by steroid intermediates from the sterol catabolic pathway, depending on the strain [147]. Alkanotrophic *Rhodococcus* strains have been shown to encode CHO involved in the conversion of β-sitosterol into stigmat-4-ene-3-one [119]. CHO of *Rhodococcus* sp. strain GK1 is induced by 4-cholestene-3-one, but not by cholesterol [147]. Induction of CHO activity in *R. erythropolis* strain ATCC25544 by cholesterol required a long incubation time (36 h), suggesting that CHO of strain ATCC25544 is not involved in the initial steps of cholesterol degradation [276]. Interestingly, three putative CHO encoding genes of *R. jostii* strain RHA1 were identified and found to be located outside of the cholesterol degradation gene cluster. These three CHO genes were not upregulated when the strain was grown on cholesterol [309]. Consistent with that, strain RHA1 growing on pyruvate and induced with cholesterol did not display any extracellular activity of CHO [250], (Chapter 2). Indeed, CHO of strain RHA1 does not catalyze the first step of cholesterol degradation. Instead, CYP125(A14) was shown to perform the initial step in strain RHA1 (Fig. 2). The role of CHO in sterol catabolism thus appears to be dependent

on the *Rhodococcus* strain. Cholesterol induced cells of strain RHA1 revealed 3βHSD activity with pregnenolone (a 3β-hydroxy steroid with a short C21 side chain), but not with cholesterol [250], (Chapter 2). By contrast, *R. rhodochrous* displayed high 3βHSD activity on cholesterol [250], (Chapter 2). Indeed, the physiological roles of CHO and 3βHSD in steroid catabolism in strains RHA1 and RG32 need further investigation.

Key enzymes involved in steroid B-ring opening: 3-ketosteroid 9α-hydroxylase

3-Ketosteroid 9α-hydroxylase (KSH) enzymes of *Rhodococcus* are known as two-component systems comprising a terminal oxygenase (KshA) and an oxygenase-reductase (KshB). KSH catalyzes an introduction of a 9α-hydroxyl group into the B ring of AD and ADD, resulting in the formation of 9α-hydroxy-4-androstene-3,17-dione (9OHAD) and 9α-hydroxy-1,4-androstadiene-3,17-dione (9OHADD), respectively [233, 304, 306] (Fig. 2). 9OHADD is chemically unstable and undergoes spontaneous transformation into a compound with an open B ring, namely 3-hydroxy-9,10-secoandrost-1,3,5(10)-triene-9,17-dione (3HSA) (Fig. 2). Substrate specificity analysis of KshA from *R. rhodochrous* strain DSM 43269 revealed that the enzyme accepts, apart from AD and ADD, also 4-androstene-17β-ol-3-one (testosterone), 4-pregnane-3,20-dione (progesterone), 19-nor-4-androstene-3,17-dione (nordion), 1-(5α)-androstene-3,17-dione (1-(5α)-AD), 5α-androstane-3,17-dione (5α-AD), 5β-androstane-3,17-dione, but not 3α-, 3β-hydroxysteroids. Steroids without a hydroxyl group at the C3 do not act as substrates [233]. Three homologous terminal oxygenase components have been identified in *R. erythropolis* strain SQ1, namely KshA1, KshA2 and KshA3 [304]. KshA1 and KshB, from strain SQ1, individually were shown to be crucial for growth on AD and ADD, as sole carbon and energy sources. Deletion of the *kshA1* gene in strain SQ1 did not result, however, in accumulation of AD and ADD when the mutant strain was grown

on phytosterols. The phytosterols were degraded by the mutant strain comparable to the wild type strain [306]. Deletion of *kshB* from strain SQ1, however, blocked sterol conversion and resulted in impaired steroid side chain degradation and the accumulation of oxidized 3-keto-stenone pathway intermediates. Interestingly, KshB from *M. tuberculosis* strain H37Rv ($KshB_{H37Rv}$) was shown to support *in vitro* activity of $CYP125A_{H37Rv}$ [27], consistent with the predicted multifunctional role of KshB proteins [306]. Conceivably, rhodococcal KshB is not only a component of the multiple KSH enzymes, but also functions as a reductase in CYP catalyzed reactions (e.g. CYP125A 26-hydroxylases) (Fig. 2).

The second terminal oxygenase homologue KshA2 was shown to possess KSH activity with AD and ADD, when incubated together with KshB. Subsequent *kshA2* promoter studies intriguingly revealed that 9OHAD, the reaction product, is a much better inducer than the substrates AD and ADD, suggesting that KshA2 might have a role in steroid catabolism different from KSH activity [304] (Fig. 2). The third terminal oxygenase KshA3 of strain SQ1 was suggested to play a role in cholesterol catabolism, displaying highest sequence identity with the homologous KshA of strain RHA1, encoded by the *ro04538* gene, located in the cholesterol degradation gene cluster [304, 309].

Additionally, the presence of 3-ketosteroid Δ^1-reductase (KSR) activity was proposed (Fig. 2), because in ADD biotransformation cultures of *R. erythropolis*, a temporal accumulation of AD was observed [304].

Key enzymes involved in steroid B-ring opening: 3-ketosteroid Δ1-dehydrogenase

3-Ketosteroid Δ^1-dehydrogenases (Δ^1 KSTDs) (EC 1.3.99.4) are flavoproteins catalyzing Δ^1-dehydrogenation steps of AD to ADD, and 9OHAD to chemically unstable 9OHADD, presumably occurring in the sterol/steroid degradation pathway (Fig. 2). *R. erythropolis*

strain SQ1 was found to produce (at least) three Δ^1 KSTDs, namely Δ^1 KSTD1, Δ^1 KSTD2 and Δ^1 KSTD3, as well as (at least) one 3-keto-5α-steroid Δ^4-dehydrogenase (Δ^4 KSTD). Activity of the latter protein was identified by native PAGE activity staining using 1-(5α)-AD as a substrate and cell free extracts of progesterone-, 4-AD- or 9OHAD-induced cells of *R. erythropolis* mutant strain RG8, with deleted genes for Δ^1 KSTD1 and Δ^1 KSTD2. Strain RHA1 possesses six putative Δ^1 KSTDs and one putative Δ^4 KSTD [142].

Δ^1 KSTD1 and Δ^1 KSTD2 from strain SQ1 are isoenzymes catalyzing conversion of AD and 9OHAD (Fig. 2) [302, 305] as well as Δ^1-dehydrogenation of 5α-androstane-3,17-dione (5α-AD), 5α-testosterone (5α-T, 17β-hydroxy-5α-androstane-3-one), progesterone, 5α-pregnane-3,20-dione (5α-P), 23,24-bisnor-5α-cholestan-3-one acid and 11β-cortisol [142]. Deletion of *kstD1*, in strain SQ1, did not affect growth of the mutant strain on AD and 9OHAD. In contrast, single *kstD2* gene deletion resulted in temporary accumulation of 9OHAD from AD [302]. A mutant strain of SQ1 with deletions in both the *kstD1* and *kstD2* genes is blocked in steroid A ring opening and stoichiometrically accumulated 9OHAD from AD [302]. Interestingly, the *kstD1 kstD2* double mutant did not accumulate any 9OHAD or AD when grown on cholesterol as sole carbon and energy source [142]. A third KSTD isoenzyme, Δ^1 KSTD3, was identified. Δ^1 KSTD3, of strain SQ1, appeared to be a unique Δ^1 KSTD, inactive with AD and 9OHAD, but active with 5α-AD (Fig. 2), 5α-T, progesterone, 5α-P; it probably utilizes atmospheric oxygen as an electron acceptor. It was speculated that 5α-T and 5α-AD, for which Δ^1 KSTD3 has high preference, are possible intermediates in the cholesterol catabolic pathway (Fig. 2) [142].

Sequencing of the region up- and downstream of the gene encoding Δ^1 KSTD3 (*kstD3*), of strain SQ1, revealed that *kstD3* is located in a cluster with genes most likely involved

in cholesterol catabolism. Genes upstream of *kstD3* encode putative 2-enoyl acyl-CoA hydratase (Hsd4B) and cholesterol oxidase (CHO); genes downstream of *kstD3* encode putative Δ^4 KSTD, HsaE (*hsaE*) and HsaG (*hsaG*). The genetic organization of this gene cluster appears to be (semi)conserved in rhodococci, such as strain SQ1, strain RHA1, *Rhodococcus ruber* strain Chol-4, as well as in several other actinomycetes [65, 142, 309].

Enzymes involved in steroid A-ring degradation: the Hsa enzymes

3-Hydroxy-9,10-secoandrost-1,3,5(10)-triene-9,17-dione (3HSA) is formed following steroid B-ring opening of AD. The 3HSA degradation pathway includes the enzymes HsaABCDEFG, found in strain RHA1 (Fig. 2 and 3), encoded by the *hsa* genes, clustering with *kstD* and *kshA* within the cholesterol catabolic gene cluster. The function of the Hsa enzymes was predicted based on significant similarity with the well studied pathway of Tes enzymes, in the Gram negative soil bacterium *C. testosteroni* strain TA441, involved in testosterone degradation. Moreover, thorough investigation of HsaAB and HsaC from strain RHA1 and *M. tuberculosis* strain H37Rv confirmed that the enzymes have a role in hydroxylation of 3HSA phenolic ring A to catechol, 3,4-DHSA, [58] and in opening of the steroid A ring via 3,4-DHSA dioxygenation to 4,9-DSHA [309, 328], respectively (Fig. 3). Furthermore, the function of HsaD from *M. tuberculosis* strain H37Rv, an ortholog of HsaD from strain RHA1, was demonstrated to be that of a 4,9-DSHA hydrolase [152, 309], consistent with its similarity with the corresponding TesD enzyme in *C. testosteroni* strain TA441 [104]. HsaD of strain RHA1 thus very likely catalyzes 4,9-DSHA hydroxylation (Fig. 3). Two products are formed upon hydrolysis of 4,9-DSHA, i.e. 9,17-dioxo-1,2,3,4,10,19-hexanorandrostan-5-oic acid (DOHNAA) and 2-hydroxyhexa-2,4-dienoic acid, which in rhodococci are further degraded by multiple

enzymatic steps (Fig. 3). Enzymes catalyzing transformation of DOHNAA remain to be characterized. HsaE probably converts 2-hydroxyhexa-2,4-dienoic acid into 4-hydroxy-2-oxohexanoic acid (Fig. 3), sharing 40% amino acid identity with TesE from *C. testosteroni* TA441 [104, 309]. It has been speculated that in *C. testosteroni* TA441 TesG transforms 4-hydroxy-2-oxohexanoic acid into pyruvic acid and propionaldehyde. TesF may further convert propionaldehyde into propionyl-coenzymeA (propionyl-CoA) [104]. HsaF is similar to TesG (48% amino acid identity) and HsaG is similar to TesF (53% amino acid identity), suggesting that similar degradation steps may take place in strain RHA1, leading to formation of pyruvic acid and propionyl-CoA, which are used for growth (Fig. 3) [309].

Regulation of steroid metabolism: KstR and KstR2 repressors

Apart from the thorough investigation of the strain RHA1 cholesterol transcriptome, identified by transcriptional profiling [309], little is known about the regulation of steroid metabolism in *Rhodococcus* strains. A putative repressor gene, encoding a TetR-type family transcriptional regulator $KstR_{SQ1}$, was identified upstream of the Δ^1 KSTD1 gene (*kstD*) of strain SQ1, suggesting $KstR_{SQ1}$ involvement in a regulatory mechanism for *kstD* transcription [307]. A transcriptional regulator (Ro04482, $KstR_{RHA1}$), with 32% similarity to $KstR_{SQ1}$, was found in strain RHA1, too. $KstR_{RHA1}$ gene is located upstream of a suite of genes within the cholesterol gene cluster predicted to be involved in the degradation of DOHNAA. Indeed, the repressor KstR of *M. tuberculosis* H37Rv (Rv3574, $KstR_{Mtb}$), orthologous to $KstR_{RHA1}$ (65% identity), was found to regulate a major portion, but not all, of the genes induced by cholesterol [134, 213]. A second TetR-type transcriptional repressor gene, encoding KstR2, has recently been identified within the cholesterol catabolic gene clusters of *R. jostii* RHA1 (Ro04598), *M. tuberculosis* H37Rv (and also of

M. smegmatis). KstR2 appears to control different subset of genes than the KstR repressor ($KstR_{RHA1}$, $KstR_{Mtb}$) does. KstR and KstR2 act independently from each other [132].

Cytochrome P450 enzymes

Cytochrome P450 monooxygenases (CYPs or P450s) form a superfamily of redox enzymes present in almost all type of organisms [211, 212], including the giant *Acanthamoeba polyphaga* mimivirus [158]. Historically "P" in the P450 name originates from the first identification of these enzymes as pigments, in 1958, in liver microsomes [80].

CYPs are *b*-type heme-thiolate enzymes. The active site commonly constitutes a non-covalently bound cofactor of an iron (III) protoporphyrin IX (heme). Ferric heme iron (Fe^{3+}) is generally characteristic for the resting state of CYP enzyme. The Fe^{3+} ion is coordinated by a cysteine via a thiolate anion (the axial ligand) and with a water molecule (acting as the distal ligand) [200, 323]. This cysteine residue is conserved in CYP proteins and typically bound *trans* to a water molecule [197]. Upon reducing a P450 enzyme and adding subsequently carbon monoxide (CO), a Fe(II)-CO complex is formed due to the *trans* bond between the cysteinate residue and CO. The Fe(II)-CO complex has a maximum in an absorption spectrum at around 450 nm, hence their designation as P450 enzymes. This is a signature feature of CYPs [295] and used to estimate CYP enzyme content [219]. Cytochrome P450 monooxygenases catalyze the incorporation of one oxygen atom into a substrate (R), resulting in formation of an oxidized product (RO), with the following general stoichiometry:

$R + O2 + NAD(P)H + H+ \rightarrow RO + NAD(P)+ + H2O$ [80].

One oxygen atom, which is incorporated into the substrate, arises commonly from a reductive cleavage, by P450s, of an atmospheric dioxygen, typically in the presence of two electrons. Electrons are transferred to P450 proteins at discrete steps in the catalytic cycle, commonly by redox partner proteins, e.g. containing flavin cofactors. Electrons are derived from hydride ion, coming from the oxidation of the reduced form of nicotinamide adenine dinucleotide (phosphate), NAD(P)H. The other oxygen atom is concomitantly reduced to water [124]. In view of their functioning in an electron transfer chain, P450 enzymes were named as "cytochromes" [256].

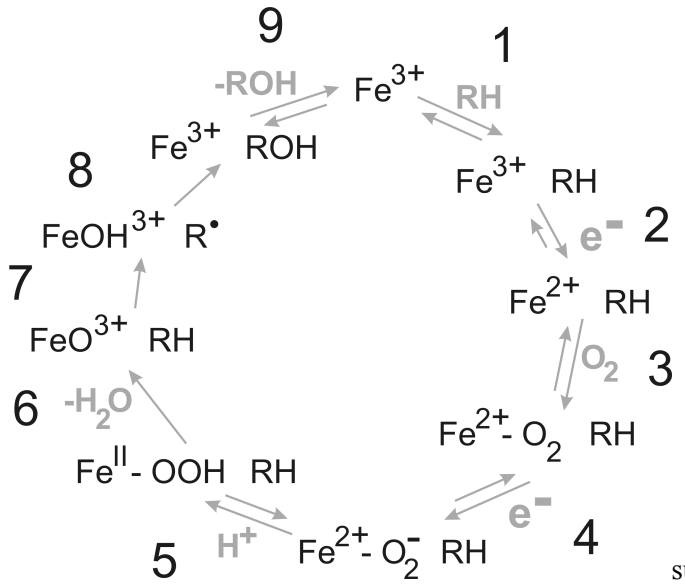


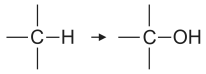
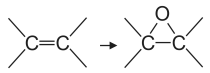
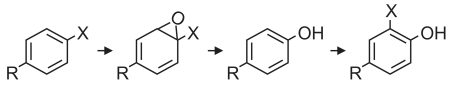
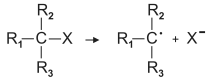
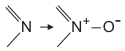
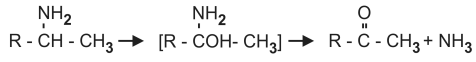
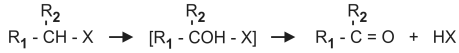
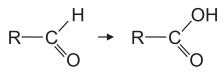
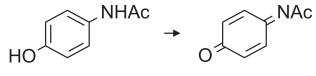
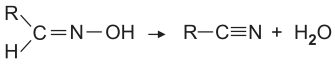
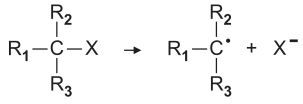
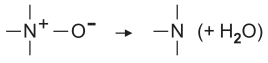
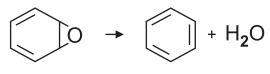
Figure 4. Simplified general catalytic cycle of P450s, with indicated steps of: (1) substrate (RH) binding to a P450 ferric form of the heme iron (Fe^{3+}) (see also text), yielding removal of the water molecule ligand and thus electron reorganizations in the heme iron d orbitals [232]; the heme iron transition takes place from a low spin state to a high spin; the reduction potential of the heme iron increases, making possible (2) a first electron reduction of the heme iron [200]; subsequently (3) oxygen binds to the reduced heme iron; (4) the oxygen complex accepts a second electron; (5) the oxygen ligand undergoes protonation and rapidly decomposes by (6) hemolytic O-O bond cleavage, yielding an active species of perferryl FeO intermediates (e.g. FeO^{3+}) and a water molecule; (7) reaction with the substrate takes place, resulting in (8) product (ROH) formation; (9) the product dissociates, resulting in restoration of the resting state (Fe^{3+}) of a P450 enzyme [114]. There are multiple equilibria present, which occur at the same time [81]. The scheme is adapted from [114] with modifications.

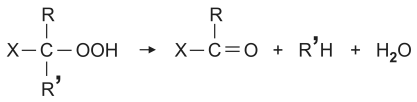
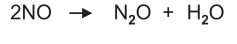
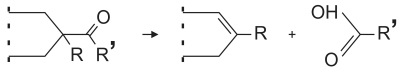
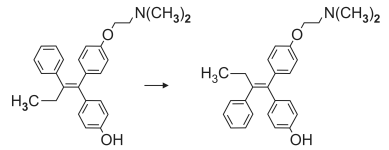
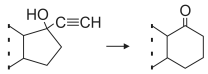
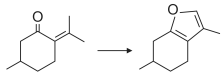
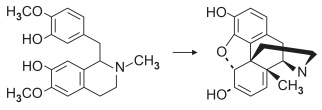
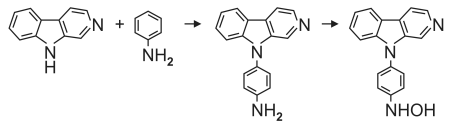
The P450 catalytic cycle of monooxygenation is very complex with transient intermediates being formed (Fig. 4) [80, 197]. Reaction intermediates are known

to collapse, e.g. when the delivery of reducing equivalents is not timely. This results in uncoupling reactions, yielding superoxide, hydrogen peroxide or water [196]. Every step of the catalytic cycle may be branched with side reactions [124], e.g. due to the inactivation of P450 enzyme into an enzyme form without the proximal ligand (P420) [200]. The nature of the catalytic cycle depends greatly on the reducing equivalent delivery cycle, as well as on the P450 enzyme and the substrate [114, 124, 232]. P450 proteins often undergo important conformational changes, in order to bind and position a substrate [80]. Some of the steps in the catalytic cycle proceed in a sequential manner, others synchronously or even in a different order [115, 124]. For instance, substrates can be bound not only by the oxidized (ferric) form of the P450 heme iron (Fig. 4), but also by the reduced (ferrous) one, e.g. the ferrous human CYP2A6 binds coumarin as a substrate [81].

P450s catalyze mostly oxygenations, but also a diversity of other reactions (Table 1), e.g. reductions, rearrangements of unstable oxygenated species, or hydroperoxides rearrangements [114]. P450s are able to transform natural compounds, e.g. steroids, plant allelochemicals, prostaglandins, thromboxanes, mammalian alkaloids, fatty acid derivatives and retinoic acid derivatives, as well as xenobiotics, e.g. drugs, pesticides, environmental pollutants, including carcinogens [43, 80, 222]. Thus, the P450 functional roles comprise xenobiotic metabolism as well as the biosynthesis of signaling molecules, critical for controlling development and homeostasis [51]. It is generally accepted that xenobiotic-metabolizing P450s are assigned to EC 1.14.14.1 and EC 1.14.13.x numbers, whereas other P450s have been assigned with other EC numbers [292].

Table 1. Examples of reactions catalyzed by cytochrome P450 enzymes, adapted from [114, 278].

Reaction name	**Reaction scheme**
Hydrocarbon hydroxylation	
Alkene epoxidation	
Arene epoxidation, aromatic hydroxylation, NIH shift	
N-dealkylation	
S-dealkylation	
O-dealkylation	
N-Hydroxylation	
N-Oxidation	
S-Oxidation	
Oxidative deamination	
Oxidative dehalogenation	
Aldehyde oxidation	
Dehydrogenation	
Dehydration	
Reductive dehalogenation	
N-Oxide reduction	
Epoxide reduction	

Reductive β-scission of alkyl peroxides	
NO reduction	
Oxidative C-C bond cleavage	
Isomerization via (abortive) oxidation	
Ring expansion and decarboxylation of an acetylene	
Ring formation	
Oxidative ring coupling	
Fusion of two substrates to form a new substrate	

Cytochrome P450 monooxygenases constitute one of the largest enzyme superfamilies known to date. [211] reported a P450 count of 11294, among which the superfamily members from plants (4266) and animals (3282) were the most abundant. P450s are classified into families, which are numbered (e.g. CYP125). Amino acid sequences identical for at least 40% are assigned to the same family. P450s sharing more than 55% of identity fall into subfamilies, which are labeled with capital letters (e.g. CYP125A). Subfamilies are divided into individual loci, labeled with numbers (e.g. CYP125A14). Furthermore, P450s that are identical for 97% or more are assigned into allelic variants, which are usually labeled with numbers (e.g. v1). The above rules for P450 nomenclature are very general and have many exceptions [256]. A P450 enzyme assignment depends

greatly on its position in the P450 phylogenetic tree. New P450s are named by the Committee on Standardized Cytochrome P450 Nomenclature [210].

P450s are the most versatile biological catalysts in nature. They are unique in their ability to efficiently transform, in various ways, unactivated substrates of a large diversity, generally hydrophobic, often in a high regio- and stereo-selective manner, typically in an aqueous environment, at neutral pH and mild temperatures, difficult for chemical synthesis [38]. Also thermostable P450s are known, which are very interesting for use in industrial processes. Harnessing P450 enzymes for industrial applications, such as production of medicines, is greatly facilitated by thorough knowledge on protein structure-function relationships. Research on P450s from humans and pathogenic microorganisms is of particular importance, providing targets for design of inhibitors and drugs [238].

P450 functional systems

At least ten classes of P450 functional systems are known (Table 2). The class I P450 system is composed of three proteins: FAD-containing ferredoxin reductase (FdR) transferring reducing equivalents from NAD(P)H to a second component, namely an iron-sulphur-containing ferredoxin (Fdx), which subsequently reduces the third component, a cytochrome P450 (Table 2). Bacterial Fdx of the class I P450 redox system were found to have redox centers of [2Fe-2S], [3Fe-4S] or [4Fe-4S]. Even systems with 7 Fe atoms, comprised of both a [3Fe-4S] and a [4Fe-4S] cluster are known [89]. The first bacterial P450 redox system was discovered in *Pseudomonas putida*; this class I system (Table 2) is composed of FAD-containing putidaredoxin reductase, [2Fe-2S]-containing putidaredoxin and camphor hydroxylase CYP101 (P450cam), [37, 128, 266].

Table 2. Classes of P450 redox systems, adopted from [89] (see text).

Class	*Occurrence*	*Redox system chain*	*Localization; remarks*	*References*
I	Most known bacteria Mitochondria of mammals, insects, nematodes	NAD(P)H→FdR→Fdx→ P450 NADH→FdR→Fdx→P450	Cytosolic; soluble Membrane associated FdR Mitochondrial matrix, soluble Fdx Inner membrane bound P450	[10, 15, 62, 131, 197]
II	Bacterium: *Streptomyces carbophilus* Microsomes	NADH→CPR→P450sca (CYP105A3) NADPH→CPR→P450 NADPH→CPR→cytb5→ P450 NADH→cytb5Red→cytb5 →P450	Cytosolic; soluble Microsomal proteins are anchored in the membrane of the endoplasmic reticulum (ER)	[253, 259]
III	Bacterium: *Citrobacter braakii*	NAD(P)H→FdR→Fldx→ P450cin	Cytosolic; soluble	[92, 138]
IV	Bacteria: *Sulfolobus tokadaii*, *S. Solfataricus*	Pyruvate→OFOR→Fdx→ P450	Cytosolic; soluble	[240, 241]
V	Bacterium: *Methylococcus capsulatus*	NADPH →FdR→Fdx-P450	Cytosolic; soluble	[120]
VI	Bacteria, e.g. *Rhodococcus rhodochrous* 11Y	NAD(P)H→FdR→Fldx-P450	Cytosolic; soluble	[6, 261]
VII	Bacteria, e.g. *Rhodococcus* sp. NCIMB 9784, *Burkholderia* sp.	NADH→PFOR-P450	Cytosolic; soluble	[197, 248]
VIII	Bacteria (e.g. *Bacillus megaterium*), Fungi (e.g. *Fusarium oxysporum*)	NADPH→CPR-P450	Cytosolic; soluble Loosely bound to the membrane	[207]
IX	Fungi (e.g. *F. oxysporum*)	NADH→P450	Cytosolic; soluble	[41]
X	Plants, mammals	P450	Bound with the membrane of the ER	[197, 229]

P450 systems of class II are mostly eukaryotic enzymes (Table 2). Class II P450 functional systems typically comprise FAD- and FMN-containing cytochrome P450 reductase (CPR) and a P450 enzyme (Table 2); an example is the CYP19 (aromatase) system, converting androgens into estrogens [129]. Other examples of P450 redox proteins of class II are: cytochrome *b*5 (cytb5) and cytochrome *b*5 reductase (cytb5Red) (Table 2). *Citrobacter braakii* uses CYP176A1 (P450cin) with a unique redox system [138]. It has FMN-containing flavodoxin (Fldx) in addition to FdR, constituting the class III P450 redox system (Table 2). 2-Oxoacid:ferredoxin oxidoreductase (OFOR), utilizing reducing equivalents from pyruvic acid (pyruvate), is a novel component found in the CYP119 redox system of e.g. *Sulfolobus solfataricus* [240], comprising class IV (Table 2). The class V P450 functional system has only a single member, found in *Methylococcus capsulatus*, namely a CYP51 protein with a C-terminally fused ferredoxin, containing [3Fe-4S], (MCCYP51FX, EC 1.14.13.70) [120]. Most likely, an unidentified separate FdR protein is also present in this system since MCCYP51FX catalysis was only possible with a surrogate reductase [120] (Table 2). In class VI P450 functional systems, the P450 protein is fused N-terminally to Fldx; its native FdR component is not known yet (Table 2). Class VII and VIII of P450 functional systems are comprised of catalytically self-sufficient proteins, namely P450-redox partner fusion proteins. In class VII, the P450 protein is fused C-terminally to an FMN-containing dioxygenase reductase domain with a [2Fe-2S]-containing ferredoxin domain (PFOR, due to the homology with phthalate-family oxygenase reductase) (Table 2). An example is CYP116B2 (P450RhF) of *Rhodococcus* sp. strain NCIMB 9784 [196, 248]. In class VIII, the P450 protein is fused N-terminally to a CPR domain (Table 2). An example is CYP102A1 (P450BM3) of *Bacillus megaterium* [77, 209]. Class IX and X of P450

systems lack P450 redox partner proteins (Table 2). CYP55A1 (P450nor, EC 1.7.99.7), from *Fusarium oxysporum* [41], represents P450 class IX [89]; it hydroxylates nitric oxide into nitrous oxide (Table 1, NO reduction), via reduction of the CYP with NADH [269] (Table 2). In class X, P450 proteins independently conduct substrate conversion via intramolecular transfer system (Table 2) [89]. An example is plant CYP74A (allene oxide synthase) participating in jasmonic acid biosynthesis by catalyzing dehydration of 13-(*S*)-hydroperoxy linolenic acid into 12,13-epoxy-linolenic acid (allene oxide), one of the key steps in the pathway [229].

While bacterial P450 enzyme systems are known to be soluble and localized in the cytoplasm, eukaryotic P450 systems are mostly bound or associated with membranes (Table 2). Eukaryotic microsomal P450s of families 1-3 are localized not only in the endoplasmic reticulum, but also in other subcellular compartments, such as plasma membranes, lysosomes and mitochondria. The physiological roles of these microsomal enzymes, present in non-endoplasmic reticulum locations, are not known [214].

Novel P450 systems continue to be discovered, especially from the analysis of genome sequences of diverse organisms. There are indications for the presence in *Pseudomonas fluorescens* of a novel P450 fused to an acyl-CoA dehydrogenase [197] and reports on a psi factor producing oxygenase (CYP6001C1) from *Aspergillus nidulans*, composed of a novel fatty acid heme peroxidase/dioxygenase domain fusion to the C-terminal P450 heme thiolate domain [21].

P450 catalysis is usually most optimal with the use of an autologous redox system. It is sometimes possible, however, to achieve P450 catalysis by exploiting heterologous redox

systems [99]. This may include construction of artificial P450-redox donor protein fusions, or use of mediators and electrodes (directly providing electrons), or the peroxide shunt [73]. In the latter case peroxides, such as hydrogen peroxide or cumene peroxide, convert ferric P450 (P450 in the resting state) into a reactive form, thus driving catalysis [149, 197].

P450s modifying steroids

Numerous native and engineered P450s catalyzing transformation of steroids are known. CYP51 is a peculiar cytochrome P450 enzyme, since it is present in all kingdoms of life. The CYP51 enzyme is an evolutionarily ancient sterol 14α-demethylase [258]. The role of CYP51 in synthesis of animal cholesterol, fungal ergosterol and plant phytosterol is well established. However, the presence of a similar CYP51 enzyme in bacteria that do not produce sterols, e.g. in *Mycobacterium* and *Rhodococcus* strains, is not yet understood. There is a strong interest in development of CYP51 inhibitors. Such compounds are potential drugs that may lower the levels of cholesterol and also act against infections by parasites (e.g. by *Trypanosoma cruzi*, by blocking the biosynthesis of the membrane sterols [32]). Alternatively, they may act as efficient fungicides and potentially also as herbicides [167].

Eukaryotic P450s with steroid transforming activities

In mammalian systems, both mitochondrial and microsomal P450 systems function in steroid converting pathways [49]. Amongst mammalian steroidogenic P450s, six CYPs (plus various hydroxysteroid dehydrogenases and 5α-reductases) are essential for biosynthesis of steroid hormones from cholesterol [18]. The cholesterol molecule is transformed first into pregnenolone by mitochondrial P450scc (CYP11A1;

EC 1.14.15.6; “scc” stands for side chain cleavage). The side chain of cholesterol is cleaved by P450scc in three consecutive hydroxylation steps [281]. The other mitochondrial steroidogenic P450s also belong to family 11, namely CYP11B1 (11β-hydroxylase, EC 1.14.15.4) and CYP11B2 (aldosterone synthase, EC 1.14.15.5), best known for catalyzing the final steps in synthesis of cortisol (glucocorticoid) and aldosterone (mineralocortiocid), respectively [24, 107]. The microsomal steroidogenic P450s are CYP17 (17α-hydroxylase/17,20-lyase), CYP21 (21-hydroxylase) and CYP19 (aromatase) [18]. CYP17 (EC 1.14.99.9) catalyzes both 17α-hydroxylation of pregnenolone and progesterone to the corresponding 17α-hydroxylated products as well as bond cleavage between C17 and C20 in 17α-hydroxypregnenolone and 17α-hydroxyprogesterone, yielding dehydroepiandrosterone and androstenedione, respectively [173]. CYP21 (EC 1.14.99.10) transforms progesterone and 17α-hydroxyprogesterone into 11-deoxycorticosterone and to 11-deoxycortisol, respectively [18], whereas aromatase converts androstenedione to estrone, testosterone to estradiol and 16α-hydroxydehydroepiandrosterone to estriol [190]. The CYP19 enzyme is unique for vertebrates in its abilities to produce estrogens from androgens in three catalytic steps, namely two hydroxylation steps at the C19-methyl and a subsequent step of the steroid A-ring aromatization [71].

Many mammalian xenobiotics metabolizing P450s also are able to modify steroids. Among these is CYP1A2, catalyzing for instance 2-, 4- and 16α-hydroxylation of 17β-estradiol [11]. CYP1A2 is the mammalian P450 that is most highly expressed in bacterial systems [137]. Moreover, CYP1A2 was shown to function with CPR, of which cofactor preference was changed from the more expensive NADPH into NADH [54].

A large group of steroid modifying CYPs is also found in fungi [23, 228, 313] and plants [216].

There is a clear interest, especially from pharmaceutical companies, in utilizing the well researched activities of mammalian P450s modifying steroids. These enzymes are, however, bound to or associated with membranes, commonly unstable and display a low catalytic activity. Optimization via enzyme engineering is thus necessary for efficient production of functional P450s in economically favored microbial hosts, instead of enzyme production by using more expensive mammalian or insect cell lines. For a long time, production of some mammalian steroid modifying P450s appeared impossible in microbial hosts [23]. However, development of efficient microbial expression systems for mammalian P450s is in progress [224]. Successful use of these P450s in microbial whole-cell biotransformations has been made possible after pathway engineering [23].

Bacterial P450s with steroid transforming activities

Soluble and typically stable bacterial P450s converting steroid molecules are an interesting alternative for the usually difficult to produce mammalian recombinant P450s. A prominent example is the formation of biologically active 1α,25-hydroxyvitamin D_3 from vitamin D_3 by CYPs from *Streptomyces* species [10, 282]. CYP105A1 from *Streptomyces griseolus* was shown to catalyze activation of vitamin D_3 in two hydroxylation steps [282]. Testosterone is a common substrate for both bacterial and mammalian CYPs (Fig. 5A, B) [3, 10, 50, 291]. Among bacterial testosterone modifying enzymes, CYP105 (P450moxA) from the actinomycete *Nonomuraea recticatena* was found to convert testosterone (Fig. 5A) into e.g. 4-androstene-3,7-dinone, in addition to the transformation of several xenobiotic compounds [10].

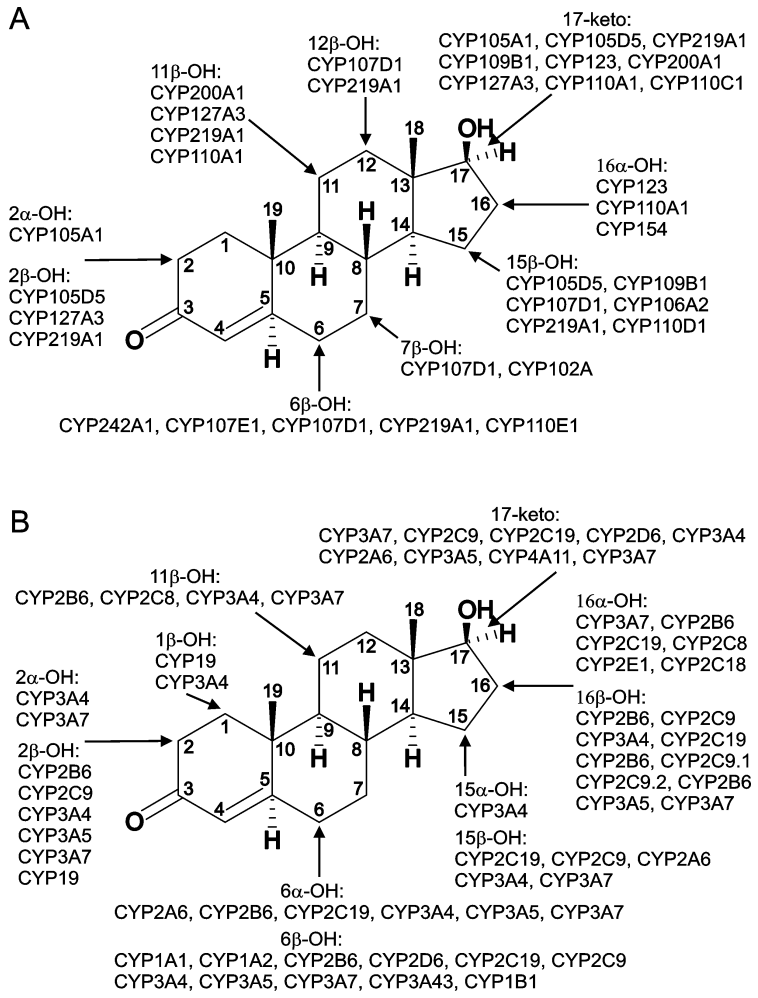


Figure 5. Testosterone carbon atom positions modified by bacterial (A) and human (B) CYPs, adapted from [10] and [3], respectively.

CYP106A2 is a native steroid modifying enzyme from *Bacillus megaterium.* CYP106A2 hydroxylates 3-oxo-Δ^4-steroids, mainly in the 15β configuration (Fig. 5A), in the presence of NADPH-dependent FdR (megaredoxin reductase) and Fdx (megaredoxin). The heterologous production yields of the enzyme in *E. coli* was optimized via directed evolution; this resulted in 1000-fold higher yields compared to CYP106A2 production in the native *B. megaterium* ATCC 13368 strain [316].

B. megaterium is the source of one of the best known and most highly active CYP, namely a self-sufficient long chain fatty acid (C12-C18) ω-1, ω-2, ω-3 hydroxylase, designated CYP102A1 or P450BM3 [77, 101, 139] (see below for more details). Rational

introduction of three amino acid mutations into the enzyme [174] resulted in construction of P450BM3tm, a triple mutant protein able to rapidly transform 4-androstene-3,17-dione into 4-androstene-3,17-dione-16β-ol (Chapter 4). Steroid hydroxylation at the 16 position by native bacterial CYPs has been observed before, but only in the α-configuration [3]. The P450BM3tm steroid 16β-hydroxylation reaction thus appears to be a unique property among bacterial steroid modifying CYPs.

Many bacteria belonging to the actinomycetes are able to grow on sterol/steroid compounds as carbon and energy sources. These bacteria, e.g. *Rhodococcus* species, encode an amazingly large number of CYP enzymes (see Chapter 5). Some of these CYP enzymes may have physiological roles in the sterol/steroid catabolic pathways. Only CYP125(A14) of *R. jostii* RHA1 has been identified as CYP enzyme involved in bacterial sterol/steroid degradation.

Flavocytochrome P450BM3 (CYP102A1)

The P450BM3 enzyme of the soil bacterium *B. megaterium* ATCC 14581 is one of the best characterized and one of the most engineered P450s to date [76], but its *in vivo* role remains to be identified [77]. Its NADPH-dependent diflavin (FAD and FMN containing) reductase domain is homologous to mammalian cytochrome P450 reductase (CPR), however, it lacks the membrane anchor typical for CPR. The CPR-like domain of P450BM3 is fused to the P450 long-chain fatty acid hydroxylase domain, forming a soluble and catalytically self-sufficient enzyme [209] (Table 2, class VIII). The two functionalities are thus present on one polypeptide chain and the P450BM3 reductase efficiently interacts with the NADPH coenzyme. Due to the fast hydride ion transfer by the reductase domain, the P450BM3 heme is reduced at rates of around 250 s^{-1} [93, 201];

it catalyzes a monooxygenation reaction with the highest known turnover rates among all P450 enzymes. Activity of P450BM3 with the most favored substrate known, namely arachidonate, exceeds 15000 min^{-1} [220]. Catalytic activity of P450BM3 depends greatly on the enzyme dimeric form, varying with altering reaction conditions, especially with differences in ionic strength and substrate concentration [141]. P450BM3 mutagenesis studies and partial crystal structures revealed the significance of specific P450BM3 enzyme residues, increasingly allowing introduction of desirable enzyme properties via enzyme engineering [75, 85, 86, 109, 136, 170, 215, 220, 221, 245, 311, 320, 321]. Ongoing research aims at solving the full-length P450BM3 crystal structure and elucidation of the enzyme catalytic process [77, 100].

Conclusions

Rhodococcus strains are catabolic marvels, increasingly used as industrial microorganisms, as attractive candidates for biotransformation and bioremediation purposes, as source and/or producers of interesting biocatalysts and various useful compounds. Among actinomycetes, rhodoccoci are easy to cultivate and efficient molecular tools are available for engineering of *Rhodococcus* strains. Research conducted on rhodococci may thus enable further harnessing of these bacteria for numerous applications, but also increase our fundamental knowledge about physiological and metabolic properties of members of the genus *Rhodococcus*, as well as related actinomycete strains.

Actinomycetes are typically rich in cytochrome P450 enzymes. Not much is however known about the *in vivo* role of these CYPs in actinomycetes, including CYPs identified

in *Rhodococcus*. Even the role of the well recognized CYP51 enzyme is not clear in actinomycetes.

Apart from CYP51, also other steroid modifying enzymes are produced by actinomycetes, especially by *Rhodococcus* strains. Rhodococci are able to grow on sterols/steroids as carbon and energy sources. At least some of these *Rhodococcus* CYPs may function in sterol/steroid catabolic pathways. Characterization of such novel rhodococcal steroid modifying enzymes may enable further engineering of the steroid degradation pathway of rhodococci, aiming to produce important bioactive steroid molecules from relatively cheap phytosterols and cholesterol. Production of useful steroids in biotransformation processes with modified *Rhodococcus* strains, avoiding chemical processes, would be advantageous for sustaining the environment.

Aim of the research

The title of my Ph.D. research project, funded by the Integration of Biosynthesis and Organic Synthesis (IBOS) NWO program of Advanced Chemical Technologies for Sustainability (ACTS), was “Expression of steroid ring modifying enzymes: selective steroid hydroxylations, steroid reductions and steroid dehydrogenations". The long term aim was to achieve production of bioactive steroid compounds from animal and plant waste material, using engineered *Rhodococcus* strains.

Wild type *Rhodococcus* strains effectively degrade sterol/steroid compounds. In previous work, Van der Geize *et al*. had constructed mutant *R. erythropolis* strains, blocked at the level of steroid ring opening, aiming at accumulating 4-androstene-3,17-dione (AD) from sterols. My research focused on the expression of selected cytochrome P450 enzymes, catalyzing modification of steroids, in such mutant *R. erythropolis* strains. Introduction of

steroid modifying P450s in such mutant *R. erythropolis* strains, incubated with sterols, might result in further conversion of AD into valuable steroid molecules.

Aromatase, the vertebrate CYP catalyzing transformation of AD to estrone, was the first enzyme that we aimed to produce in these mutant *R. erythropolis* strains. Despite many efforts, all attempts to achieve sufficient expression of aromatase in *R. erythropolis* failed (data not shown; see Chapter 5).

Subsequently, we observed that heterologous expression of an engineered bacterial CYP, namely P450BM3tm of *B. megaterium*, was possible in these mutant *R. erythropolis* strains, also resulting in efficient transformation of AD (Chapter 4). In further work we also succeeded in expression of various *R. jostii* P450 enzymes in the mutant *R. erythropolis* strains, blocked in steroid ring degradation.

Previously, Van der Geize *et al*. had characterized various rhodococcal and mycobacterial KSTD and Ksh enzymes involved in steroid degradation. In further work, we aimed to completely elucidate the bacterial sterol/steroid catabolic pathway and to characterize further enzymes involved. A thorough investigation of the native set of steroid catabolic enzymes also would provide a firm basis for the design and construction of mutant strains efficiently producing bioactive steroids. The availability of the *R. jostii* strain RHA1 genome sequence allowed a detailed transcriptomic analysis. Investigation of *R. jostii* strain RHA1 cells grown on cholesterol revealed numerous upregulated genes, including various *CYP*s [309]. Possible roles for such CYPs in rhodococcal steroid catabolism remained to be identified. Six *CYP*s upregulated on cholesterol were chosen for expression in *R. erythropolis*. Three of these *R. jostii* strain RHA1 CYPs, namely CYP125(A14), CYP257A1 and CYP51, were successfully produced in *R. erythropolis* strains, in relatively high expression levels and in active and soluble forms, as judged

from CO-difference spectra. CYP51 is a well studied enzyme in animals, plants and fungi. The physiological role and substrate specificity of rhodococcal CYP51s had not been characterized yet. Unfortunately, lack of availability of known CYP51 substrates hampered further analysis of this *R. jostii* strain RHA1 protein. The more detailed analysis of the RHA1 P450 proteins, namely CYP125(A14) and CYP257A1 is described in this thesis (Chapter 2 and 3, respectively).

Outline of this thesis

Chapter 1 provides a broad overview on rhodococci, steroids and cytochrome P450 enzymes, serving as an introduction for this thesis.

Chapter 2 describes the functional role of the CYP125 enzyme in sterol degradation by *Rhodoccocus* strains. The enzyme of *R. jostii* strain RHA1, designated CYP125(A14), is proven to be crucial for initiating the degradation of 3-hydroxysterols. The corresponding CYP enzyme of *R. rhodochrous* DSM43269 mutant strain RG32 ($CYP125_{DSM43269}$) is shown to be necessary for 3-hydroxysterol side-chain degradation. Differing from CYP125(A14), $CYP125_{DSM43269}$ does not initiate the 3-hydroxysterol degradation pathway. Different 3-hydroxysterol catabolic pathways, both using the same sterol C26-hydroxylase activity of CYP125, were proposed to occur in these two *Rhodococcus* species.

Chapter 3 describes characteristics of the *R. jostii* strain RHA1 CYP257A1 enzyme, its ability to catalyze N-demethylation of dextromethorphan into 3-methoxymorphinan and its putative physiological role in sterol metabolism.

Chapter 4 exemplifies *Rhodococcus* as an interesting and cost-efficient host for the heterologous expression of active P450BM3tm enzyme, able to perform steroid

bioconversions. The stereo- and regio-selectivity and activity of P450BM3tm towards 4-androstene-3,17-dione was determined as 16β-hydroxylase.

Chapter 5 gives the summary of this thesis, a discussion of the results, reported in the experimental chapters, and an outlook for future research.